

REMARKS

The present application was filed on October 31, 2000 with claims 1-27. In the outstanding Office Action, the Examiner: rejects claims 1-27 under 35 U.S.C. §102(e) as being unpatentable over U.S. Patent No. 6,112,203 to Bharat et al. (hereinafter "Bharat").

In this response, Applicants amend independent claims 1, 10 and 19, and traverse the §102(e) rejection for at least the following reasons.

Regarding the §102 rejection of claims 1-27 based on Bharat, Applicants respectfully assert that Bharat fails to teach or suggest all of the limitations of claims 1-27.

While Applicants believe that the claims of the present application in their form prior to this Amendment were patentably distinguishable over Bharat for at least the reasons given in their Appeal Brief dated March 28, 2005, Applicants have nonetheless amended the independent claims in a sincere effort to expedite the present application through to issuance.

Accordingly, the present invention, for example, as recited in amended independent claim 1, provides a computer-based method of performing document retrieval in accordance with an information network. The method comprises the steps of retrieving one or more documents from the information network that satisfy a user-defined predicate, collecting statistical information about the one or more retrieved documents as the one or more retrieved documents are analyzed, and using the collected statistical information to automatically determine further document retrieval operations, such that one or more documents are retrieved from the information network that satisfy the user-defined predicate without assuming a predefined link structure. The newly added language is underlined for emphasis. Independent claims 10 and 19 have been amended in similar manner.

As illustratively explained in the present specification at page 4, line 22, through page 5, line 20:

The present invention provides a more interesting and significantly more general alternative to conventional crawling techniques. As is evident from the teachings herein, no specific model for web linkage structure is assumed in intelligent crawling according to the invention. Rather, the crawler gradually learns the linkage structure statistically as it progresses. By linkage structure, we refer to the fact that there is a certain relationship between the content of a web page and the candidates that it links to. For example, a web page containing the word "Edmund Guide" is likely to link to web pages on automobile dealers. In general, linkage structure refers to the relationship between the various features

of a web page such as content, tokens in Universal Resource Locators (URL), etc. Further, in general, it is preferred that the linkage structure be predicate-dependent. An intelligent crawler according to the invention learns about the linking structure during the crawl and find the most relevant pages. Initially, the crawler behavior is as random as a general crawler but it then gradually starts auto-focusing as it encounters documents which satisfy the predicate. A certain level of supervision in terms of documents which satisfy the predicate may be preferred since it would be very helpful in speeding up the process (especially for very specific predicates), but is not essential for the framework of the invention. This predicate may be a decision predicate or a quantitative predicate which assigns a certain level of priority to the search.

The intelligent crawler of the invention may preferably be implemented as a graph search algorithm which works by treating web pages as nodes and links as edges. The crawler keeps track of the nodes which it has already visited, and for each node, it decides the priority in which it visits based on its understanding of which nodes is likely to satisfy the predicate. Thus, at each point the crawler maintains candidate nodes which it is likely to crawl and keeps re-adjusting the priority of these nodes as its information about linkage structure increases (Underlining added for emphasis).

In contrast, Bharat discloses a method for ranking documents in a hyperlinked environment using connectivity and content analysis (see Abstract of Bharat). Thus, Bharat does not teach or suggest “using the collected statistical information to automatically determine further document retrieval operations, such that one or more documents are retrieved from the information network that satisfy the user-defined predicate without assuming a predefined link structure,” as recited in the claimed invention.

As is made clear at columns 2 and 3, the techniques of Bharat require a preconstructed graph that includes nodes and directed edges, where each node represents a document and the directed edges represent the links connecting the documents. This is what Bharat refers to as a “start-set.” Thus, Bharat assumes a predefined link structure.

However, while the method of the invention may generate a graph-like structure as it “crawls,” it merely starts off with a start list which is merely a list of Uniform Resource Locators (URLs), see, for example, page 8 and 9 of the present specification. As explained above, the inventive method keeps track of the nodes which it has already visited, and for each node, it decides the priority in which it visits based on its understanding of which nodes is likely to satisfy the predicate. Thus, at each point, the inventive method maintains candidate nodes which it is likely to crawl and keeps re-adjusting the priority of these nodes as its information about linkage structure

increases. This is an illustrative example of what is meant by “using the collected statistical information to automatically determine further document retrieval operations, such that one or more documents are retrieved from the information network that satisfy the user-defined predicate without assuming a predefined link structure,” as recited in the claimed invention. Bharat does not teach or suggest this limitation.

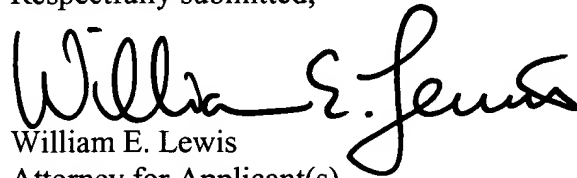
Applicants also assert that Bharat does not disclose the step of “collecting statistical information about the one or more retrieved documents as the one or more retrieved documents are analyzed,” as in the claimed invention. While Bharat does disclose content analysis, it does not appear that any “statistical information” is being collected in the Bharat document ranking technique.

For at least the above reasons, Applicants respectfully assert that independent claims 1, 10 and 19 are patentable over Bharat.

The remainder of the claims (namely, claims 2-9, 11-18 and 20-27) rejected over Bharat depend, either directly or indirectly, from claims 1, 10 or 19, which are believed patentable for the reasons set forth above. Furthermore, the remaining claims define additional patentable subject matter in their own right.

In view of the above, Applicants believe that claims 1-27 are in condition for allowance, and respectfully request withdrawal of the §102(e) rejections.

Respectfully submitted,



William E. Lewis
Attorney for Applicant(s)
Reg. No. 39,274
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-2946

Date: September 26, 2005